N94-24190

THE USE OF ANALYTICAL MODELS IN HUMAN-COMPUTER INTERFACE DESIGN

Leo Gugerty
Lockheed Engineering and Sciences Company

Research directed by Marianne Rudisill, Manager, Human Computer Interaction Lab, NASA JSC.

INTRODUCTION

Researchers in the human-computer interaction (HCI) field commonly advise interface designers to "know the user." Various approaches are currently used to get information about the user into the hands (and mind) of the designer. One approach is to use design guidelines (e.g., NASA Johnson Space Center, 1988) which can incorporate knowledge of human psychological strengths and weaknesses and make them accessible to designers. However, guidelines give only overview information. They do not help the designer to configure the interface for a specific task and specific users (Gould and Lewis, 1985). Another way to know the user is to conduct usability tests (Gould and Lewis, 1985). This involves building prototype interfaces as early as possible in the design process, observing typical users as they work with the prototype, and fixing any observed problems during the next iteration of the design. While effective in making the designer aware of user needs, usability testing adds a significant amount of time to the design of user interfaces.

Recently, a large number of HCI researchers have investigated another way to know the user - building analytical models of the user, which are often implemented as computer models. These models simulate the cognitive processes and task knowledge of the user in ways that allow a researcher or designer to estimate various aspects of an interface's usability, such as when user errors are likely to occur. This information can lead to design improvements. Analytical models can supplement design guidelines by providing designers rigorous ways of analyzing the information-processing requirements of specific tasks (i.e., task analysis). These models offer the potential of improving early designs and replacing some of the early phases of usability testing, thus reducing the cost of interface design.

This paper describes some of the many analytical models that are currently being developed and evaluates the usefulness of analytical models for human-computer interface design. The paper is intended for researchers who are interested in applying models to design and for interface designers. This is a summary of an extensive literature review paper on the use of analytical models in design that is being conducted at the Johnson Space Center's Human-Computer Interaction Laboratory.

The question of whether analytical models can really help interface designers is currently receiving much attention in the field of human-computer interaction. Advocates of model-based design claim that our knowledge of cognitive psychology is becoming sophisticated enough to allow analytical models of the user to play a useful role in interface design (Kieras, 1988; Butler, Bennett, Polson, and Karat, 1989). Modeling proponents suggest that models could be used during interface design in two important ways:

1. Models can help designers conduct a rigorous task analysis, which in turn may help generate design ideas. A number of analytical models (e.g., the GOMS model, Card, Moran, and Newell, 1983) involve specifying the goals, actions, and information requirements of the user's task. Research suggests that these task

analyses can help designers generate effective design ideas.

2. After interface designs have been generated, models can help evaluate their effectiveness. A human-factors psychologist or engineer could work with a designer to build a computer model of how a user would interact with a new interface. This model could be run with various input conditions to predict how long the user will take to perform tasks using the interface, and likely sources of user errors.

The benefits of analytical models are by no means universally accepted in the HCI community. Many HCI researchers and practitioners have questioned the usefulness of models for interface design. Whiteside and Wixon (1987) claim that current models are only applicable to the specific task and context for which they were developed and cannot be applied to new interfaces. Others (e.g., Curtis, Krasner, and Iscoe, 1988; Rossen, Maas, and Kellogg, 1988) suggest that models may not fit in with the needs of design organizations or with the intuitive thinking and informal planning that designers sometimes use.

This paper will focus on computational, analytical models, such as the GOMS model, rather than less formal, verbal models, because the more exact predictions and task descriptions of computational models may be useful to designers. The literature review paper that is summarized here evaluated a number of models in detail, focusing on the empirical evidence for the validity of the models. Empirical validation is important because without it models will not have the credibility to be accepted by design organizations. This paper will briefly describe two analytical models in order to illustrate important conclusions from the literature review. Following this, the paper will discuss some of the practical requirements for using analytical models in complex design organizations such as NASA.

EMPIRICAL EVALUATION OF ILLUSTRATIVE MODELS

GOMS MODEL

The GOMS model was developed as an engineering model to be used by HCI designers, and it has received much more empirical testing than any other analytical model of HCI tasks. Many of the issues concerning the use of GOMS models in design are relevant to other analytical models as well.

GOMS models are applicable to routine cognitive skills. They are best suited for tasks where users make few errors. More open-ended tasks that involve extensive problem solving and frequent user errors (e.g., troubleshooting) are not good candidates for GOMS modeling.

GOMS stands for goals, operators, methods, and selection rules, the four elements of the model. GOMS models are hierarchical. The assumption is that at the highest level people's behavior on a routine computer task can be described by a hierarchy of goals and subgoals. At the most detailed level, behavior is described by operators, which can be physical (such as typing) or mental (such as comparing two words). Operators that are often used together as a unit are built up into methods. For example, one might have a standard method of deleting text in a text editor. Sometimes more than one method can meet a goal and selection rules are used to choose among them.

GOMS models can help an interface designer get a qualitative understanding of the goal structure and information requirements of a task (i.e., a task analysis). In addition, Kieras and Polson (1985) developed a formal implementation of GOMS models, Cognitive Complexity Theory (CCT), that allows designers to make quantitative statements about users' errors, learning time, and performance time for particular interfaces. In CCT, GOMS models are represented as production systems. In a production system the parts of a GOMS model are represented by a series of if-then rules (production rules) that can be run as a computer simulation model. A number of quantitative metrics can be derived from a CCT production system that, according to proponents of CCT, can be used to predict users' performance on a task (Kieras, 1988; Olson and Olson, in press). For example, task learning time, task performance time, and the number of user errors can be predicted.

To date, GOMS models have not been used to help design a commercial interface. Most empirical studies of GOMS models have been evaluations of existing interfaces that were designed without using GOMS. For example, Bovair, Kieras, and Polson (in press) evaluated GOMS estimates of task performance time for existing interfaces. Using a text editing task, they found that the number of production-system cycles and of certain complex operators (such as looking at the text manuscript) could match performance time fairly well, explaining about 80% of the variability of users' performance times across editing tasks.

It is important to point out that in studies like this data (such as errors and the time to learn and perform tasks) are collected from users of an interface, and statistical techniques (such as regression) are used to determine whether the GOMS predictions match the data. In these studies, GOMS models are not used to make *a priori* predictions of user performance. Rather, the models' estimates of user performance are statistically compared to the empirical data to see how much of the variability in users' performance data can be explained by the model. Although some researchers suggest that GOMS models can be used to make a priori predictions of user performance (Olson and Olson, in press), this has not been done successfully to date.

In addition to evaluations of existing interfaces, a few studies have looked at how GOMS models can be used to generate ideas for redesigning interfaces. These studies take advantage of the fact that GOMS models provide a detailed task analysis (i.e., a representation of the goals, subgoals, and procedural steps) required to perform a task. Elkerton and Palmiter (1989) used a GOMS model of the knowledge required for Hypercard authoring tasks to design a menu-based Hypercard help system that allowed faster information retrieval and that was liked better than the original help system.

This study is important because it shows that GOMS models can be used for more than post-hoc evaluation of existing designs. In this study, the task analyses provided by GOMS models were used to generate computer-related artifacts (in this case, procedural instructions). In addition, these artifacts were generated fairly directly from the task analyses without extensive interpretation or "judgment calls."

To summarize the empirical evaluation of GOMS models, models developed for a single, existing interface can be used in a *post-hoc*, quantitative fashion to explain performance time, learning time, and number of errors with that interface. No one has yet tested whether GOMS models can make accurate quantitative performance predictions for an interface that is still in design. However, encouraging progress has been made in using the task analyses provided by a GOMS model to help generate effective instructions that can be incorporated in help systems and user manuals.

TULLIS' MODEL

The next model to be described has a much narrower range of application than GOMS models and focuses on general psychological processes rather than task analysis. Perhaps because of these differences, this model, developed by Tullis (1984), is better than GOMS at making *a priori* predictions of user performance. Tullis' model focuses on aspects of a display, such as display density, that affect how well people can find information in the display. It emphasizes general processes, such as perceptual grouping, that affect display perception regardless of the content of the display. The effects of task knowledge on display perception (e.g., effects of user expertise) are not considered. Tullis' model is applicable only to alphanumeric displays that make no use of color or highlighting. The model has been applied to simple search tasks involving displays for airline and motel reservations and for aerospace and military applications (Tullis, 1984).

Based on a literature review, Tullis hypothesized that five factors would affect the usability of alphanumeric displays: overall density, local density, number and size of the perceptual groups, and layout complexity. He developed operational definitions so that quantitative values could be calculated for each factor, given a display layout as input. Then, he conducted an experiment in which subjects searched for information in displays and rated the usefulness of the displays. Regression analyses showed that the five factors could explain subjects' search times and subjective ratings fairly well.

Tullis implemented his regression model in the Display Analysis Program (Tullis, 1986). This program accepts a display layout as input. It outputs quantitative estimates of overall density, local density, number of perceptual groups, and average group size. It also provides graphical output describing the display density analysis and the perceptual groups. Finally, it predicts average search time and subjective ratings for the display.

Tullis (1984) then used his model to predict search times and subjective ratings for a second experiment, using different subjects and displays than the experiment that was used to develop the regression equations. The predicted search times and subjective ratings matched the actual times and ratings fairly well, with a correlation of about 0.64 (r^2) for each variable. The model correctly predicted the displays with the best search time and rating. Tullis' model was also able to predict search times from three previous studies in the literature ($r^2 > 0.63$ in each study) (Tullis, 1984). However, when Tullis' model was tested on tasks more complex than simple display search, it did not predict subjects' performance well (Schwartz, 1988).

To summarize, Tullis' model is applicable within a limited domain—inexperienced users performing simple search tasks involving alphanumeric displays. Within this domain, however, the model's performance is impressive. Tullis has taken the step that GOMS users have neglected and used his model to predict performance for displays and subjects different from the ones on which the model was developed. The model was able to predict well in these cases. One disadvantage of Tullis' model is that it neglects cognitive factors affecting display perception, such as the effect of a user's task knowledge.

CONCLUSION: EMPIRICAL EVALUATION OF ANALYTICAL MODELS

Earlier in the paper, it was suggested that analytical models could be used in interface design in two ways. The first of these involves using models early in the design process to conduct rigorous task analyses, which are then used to generate ideas for preliminary designs (e.g., menu structures). The second potential use of models occurs later in the design process, after preliminary designs have been developed. In this case models are used to evaluate designs by making quantitative predictions about expected user performance given a particular design.

The empirical evidence considered in the literature review, and summarized here, suggests that, except for one model with a narrow range of application, there is no empirical evidence that analytical models can predict user performance on a new interface. There is some encouraging evidence that analytic models used for task analysis can help in the process of generating designs; however, this conclusion is based on only a few studies. The review of the empirical evidence suggests, then, that future research aimed at demonstrating model-based improvements in interfaces should focus on three areas:

- Replicating and extending the studies of model-based interface redesign (e.g., Elkerton and Palmiter, 1989).
- Demonstrating model-based interface design for a new interface.

- Demonstrating the predictive use of models to evaluate preliminary designs.

Based on the empirical evidence to date, the first two of these would be the most promising avenues of research.

What are some possible reasons for the failure of models to accurately predict performance with a new interface? It may be that critics such as Whiteside and Wixon (1987) are correct in that people's procedures, goals, and cognitive operators are too context specific to allow prediction in a context as different as a new interface. A large body of research in cognitive psychology suggests that experts' performance in a particular domain is largely dependent on domain-specific knowledge, as opposed to general-purpose cognitive skills (Chi, Glaser, and Rees, 1982; Glaser, 1984). And models such as GOMS focus primarily on the task-specific knowledge of experienced users. It is interesting that the model that was able to predict user performance on a slightly different interface (Tullis') is not a task analytic model. Tullis' model focuses on general perceptual abilities. This suggests that in order to predict performance for new interfaces, task analytic models must include more explicit representation of how general purpose cognitive characteristics (such as working memory limitations) affect user performance.

An addition should be made to the above list of research areas. This suggestion is based on the fact that there are no empirically validated models that can describe HCI tasks involving higher-level cognitive processes such as problem solving. However, space-related computer systems are rapidly becoming intelligent enough to assist people in complex tasks, such as medical diagnosis and scientific research, which involve more complex cognition. Models are currently being developed with the goal of describing these more complex tasks in a way that is useful to interface designers. An example is the Programmable User Models (PUMs) (Young and Whittington, 1990). However, most of these models have not been empirically validated.

A fourth area of further research, then, is:

- Developing and testing models of complex HCI tasks involving high-level cognitive processes.

USING MODELS IN DESIGN ORGANIZATIONS

So far, this paper has focused on whether analytical models can improve interface designs. However, even if models were conclusively demonstrated to improve interfaces, this would still not ensure their use by design organizations such as NASA. What is needed is evidence for the usefulness as well as the validity of models. That is, it must be shown that models can meet the needs of individual designers (e.g., preferred design methods) and of design organizations (e.g., cost, scheduling, and personnel constraints).

With respect to individual designers, an understanding of the various ways that designers generate, develop, and evaluate ideas is needed. Analytical models would be provided to designers as detailed procedures or as software tools. The principle of considering the cognitive and motivational processes of users applies to model developers just as it does to the designers of other software tools. In short, designers are users too. Therefore, if model developers want their models to be used in actual design projects, they must either construct their models to fit in with the preferred design processes of designers or provide ways of training designers to use the models.

But decisions regarding the commercial use of models are made by managers, not by individual designers. Therefore, models also must be shown to meet the multifaceted needs of design organizations, for example, cost, schedule, and personnel requirements. This section will discuss the problems that must be overcome before analytical models are accepted by designers and their work organizations.

NEEDS OF INDIVIDUAL DESIGNERS

Two studies conducted by Curtis and his colleagues showed that major difficulties in software design are caused by a lack of application-domain knowledge on the part of designers. (Curtis et al., 1988; Guindon, Krasner, and Curtis, 1987). The analogous problem in the case of interface design would be a lack of knowledge of the user's task. When Rosson et al., (1988) interviewed interface designers about the techniques they used to generate design ideas, they found that the most frequently mentioned techniques (about 30%) were for analyzing the user's task. Most of this task analysis involved informal techniques, such as interviewing users or generating a task scenario.

These findings present both an opportunity and an obstacle to the use of models by interface designers. First, since designers often lack knowledge of the user's task and spend a large amount of effort getting it, they might see the usefulness of task analytic models such as GOMS. The potential obstacle is that designers may prefer to stick with their informal techniques, instead of the more rigorous task analytic models. Rosson et al., suggest that tools to aid in idea generation should primarily support designers' informal techniques. Lewis, Polson, Wharton, and Rieman (1990) offer an interesting way of combining formal modeling with a technique currently used by software designers—design walkthroughs. They developed a formal model of initial learning and problem solving in HCI tasks, and then derived from the model a set of structured questions (a cognitive walkthrough) that can be used to evaluate the usability of an interface.

This discussion presents only an example of the kind of issues that need to be considered regarding the needs of individual designers. Further research is needed on the cognitive and motivational processes of designers and what these processes suggest about the design of analytic models.

NEEDS OF DESIGN ORGANIZATIONS

The Curtis et al., (1988) study mentioned above also considered the organizational aspects of software design. In addition, Grudin and Poltrock (1989) conducted an extensive interview study of the organizational factors affecting interface design. Some of the findings of these studies that relate to the use of analytical models are discussed below.

An important characteristic of many computer-system design organizations is complexity. Many groups may contribute to a final design product: interface and system designers, human factors personnel, training developers, technical writers, and users (e.g., astronauts). Curtis et al., (1988) noted a wide variety of communications problems that resulted because of this organizational complexity. One such problem arises when groups interpret shared information differently because of differences in background knowledge. This could easily cause problems, for example, if the people in an organization who are experienced with modeling (e.g., a designer or human factors expert) have to communicate the results of a modeling analysis to a project manager. A possible solution to this problem of misinterpretation is for model developers to make the structure and outputs of their models as clear as possible.

In addition to communication problems, another problem arising from the variety of roles in design organizations has to do with personnel and training. A manager considering the use of models on a design project faces a number of questions along these lines. Can existing personnel do the modeling (e.g., designers or human factors personnel)? How much training will they require? If new personnel must be hired, what kinds of background must they have? Model developers must have answers to these questions.

One answer comes from the work of Kieras (1988). He has developed and published a procedure for building GOMS models. Informal testing showed that computer science undergraduates could use this procedure to generate GOMS models and make usability predictions "with reasonable facility." More than this is necessary, however. Validation studies must be done to test whether the personnel that would use models in design organizations can build models that make the same kinds of predictions as the experts who initially developed the model. These studies should also document the kind of training necessary to achieve these ends.

In addition to complexity, other characteristics of design organizations that affect their openness to modeling are strict project scheduling and a concern with monetary costs. Detailed estimates are needed of the time and money costs of using analytical models in commercial design.

CONCLUSION: THE USE OF ANALYTICAL MODELS IN INTERFACE DESIGN

Can the use of analytical models be recommended to interface designers? Based on the empirical research summarized here, the answer is: not at this time. There are too many unanswered questions concerning the validity of models and their ability to meet the practical needs of design organizations. However, some of the research described here suggests that models can be of practical use to designers in the near future. Of special interest is the research that used models as task analytic tools to generate interface design ideas (e.g., Elkerton and Palmiter, 1989).

This paper has suggested research and development that is necessary in order for analytical models to be accepted by complex design organizations. These suggestions are summarized in Table 1. It seems that the empirical research on analytical models gives good reason to pursue the research and development goals outlined here.

TABLE 1.

Methods of Increasing the Use of Analytical Models in Interface Design

Demonstrate design improvements:
- Validate model-based interface redesign.
- Validate model-based interface design.
- Validate predictive use of models to evaluate preliminary designs.
- Develop and validate models of complex HCI tasks involving high-level cognitive processes.

Meet the needs of individual designers:
- Study the design methods and cognitive processes of individual designers.
- Change the models and/or develop training materials to ensure that models fit in with designers' methods and cognitive processes.

Meet the needs of design organizations:
- Make models' structure and outputs easily interpretable.
- Develop means of training designers to use models. Validate that this training works and document the costs of training.
- Document the time and monetary costs of using models.

ANALYTICAL MODELS AND SPACE-RELATED INTERFACE DESIGN

So far, this paper has provided a general analysis of the use of analytical models in human-computer interface design. How much of this analysis is applicable to the design of space-related interfaces? The Human-Computer Interaction Laboratory (HCIL) at the Johnson Space Center is currently conducting preliminary task analyses for the tasks required on a long-duration space mission, such as a mission to Mars (Gugerty and Murthy, in preparation). This work suggests that the range of tasks on such a mission is quite broad—ranging from reading to controlling complex equipment to conducting scientific research. The possible information technologies for long-term missions are also quite diverse, for example, workstations for supervisory control, graphics workstations for scientific research, computer-supported group meetings, medical expert systems, and virtual workstations for telerobotic control. It seems that space-related tasks are diverse enough to span almost the entire range of human-computer interaction tasks. Therefore, the general analysis of this paper will be applicable to space-related tasks in most cases.

One project in the JSC HCIL is focusing on the use of analytical models in designing medical decision support systems for space crews. This project is following up on the work of Elkerton and Palmiter (1989) in which GOMS was used as a task analytic model to help generate interface design ideas. One medical task that space crew members will face is learning or relearning medical procedures from computer displays. This project will test whether building GOMS models of medical procedures can help interface designers build better interfaces for displaying this procedural information. The GOMS approach will be compared with other methods of task analysis, including psychological scaling techniques such as the Pathfinder algorithm (McDonald and Schvaneveldt, 1988).

REFERENCES

1. Bovair, S., Kieras, D. E., and Polson, P. G. The acquisition and performance of text editing skill: A production system analysis, *Human Computer Interaction*, in press.

2. Butler, K, Bennett, J., Polson, P., and Karat, J. Report on the workshop on analytical models, *SIGCHI Bulletin*, 20, 4, 1989, pp. 63-79.

3. Card, S. K., Moran, T., and Newell, A. *The psychology of human-computer interaction*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1983.

4. Chi, M. T. H., Glaser, R., and Rees, E. Expertise in problem solving, in R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1), Lawrence Erlbaum Associates, Hillsdale, NJ, 1982.

5. Curtis, B., Krasner, H., and Iscoe, N. A field study of the software design process for large systems, *Communications of the ACM,* 31,11, 1988, pp. 1268-1287.

6. Elkerton, J. and Palmiter, S., Designing help systems using the GOMS model: An information retrieval evaluation, *Proceedings of the Human-Factors Society 33rd Annual Meeting*, Santa Monica, CA, 1989.

7. Glaser, R. Education and thinking: The role of knowledge, *American Psychologist*, 39, 2, 1984, pp. 93-104.

8. Gould, J. D. and Lewis, C. H. Designing for usability and what designers think, *Communications of the ACM*, 28, 3, 1985, pp. 300-311.

9. Grudin, J. and Poltrock, S. E. User interface design in large corporations: Coordination and communication across disciplines. In K. Bice and C Lewis (Eds.), *Proceedings of the CHI 89 Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY 1989.

10. Gugerty, L. J. and Murthy, P. (in preparation). Using task analysis to guide information-technology development for future space missions, Paper to be presented at the AIAA Space Programs and Technologies Conference, September, 1990.

11. Guindon, R., Krasner, H., and Curtis, B. Breakdowns and processes during the early activities of software design by professionals. In G. M. Olson, S. Sheppard, and E. Soloway (Eds.), *Empirical studies of programmers: Second workshop*, Ablex Publishing Corp., Norwood, NJ, 1987.

12. Kieras, D. E. Towards a practical GOMS model methodology for user

interface design. In M. Helander (Ed.), *Handbook of human-computer interaction,* Elsevier Science Publishers, Amsterdam, 1988.

13. Kieras, D. E. and Polson, P G. An approach to the formal analysis of user complexity, *International Journal of Man-Machine Studies,* 22, 1985, pp. 365-394.

14. Lewis, C., Polson, P., Wharton, C., and Rieman, J. Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces, in J. C. Chew and J. Whiteside (Eds.), *Proceedings of the CHI 90 Conference on Human Factors in Computing. Systems*, Association for Computing Machinery, New York, NY, 1990.

15. McDonald, J. E. and Schvaneveldt, R. W. The application of user knowledge to interface design. In R. Guindon (Ed.), *Cognitive science and its applications for human-computer interaction*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.

16. NASA Johnson Space Center, Space Station Freedom Program human-computer interface guide (USE 1000, Version 2.1), Johnson Space Center, Houston, TX, 1988.

17. Olson, J. R. and Olson, G. M. The growth of cognitive modeling in human-computer interaction since GOMS, *Human-Computer Interaction*, in press.

18. Poltrock, S. E. Innovation in user interface development: Obstacles and opportunities. In K. Bice and C. Lewis (Eds.), *Proceedings of the CHI 89 Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, 1989.

19. Rosson, M. B., Maass, S., and Kellogg, W. A. The designer as user: Building requirements for design tools from design practice, *Communications of the ACM,* 31, 11, 1988, pp. 1288-1299.

20. Schwartz, D. R. The impact of task characteristics on display format effects, *Proceedings of the Human Factors Society - 32nd Annual Meeting*, Human Factors Society, Santa Monica, CA, 1988.

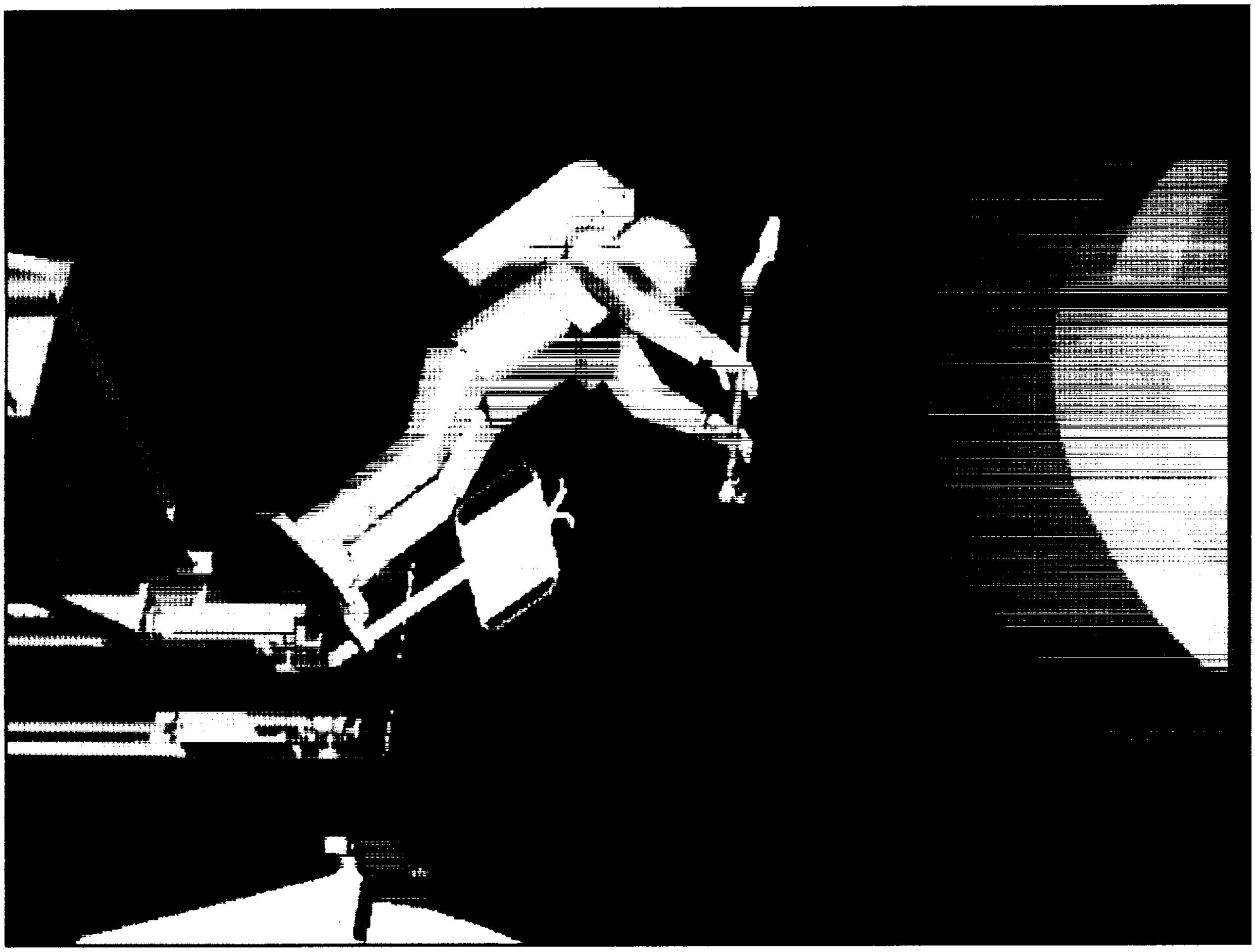
21. Tullis, T. S. Predicting the usability of alphanumeric displays, Doctoral Dissertation, Rice University, The Report Store, Lawrence, Kansas, 1984.

22. Tullis, T. S. A system for evaluating screen formats, *Proceedings of the Human-Factors Society 30th Annual Meeting*, Human Factors Society, Santa Monica, CA, 1986.

23. Whiteside, J. and Wixon, D. Discussion: Improving human-computer interaction - A quest for cognitive science. In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human-computer interaction*, MIT Press, Cambridge, MA, 1987.

24. Young, R. M. and Whittington, J. Using a knowledge analysis to predict conceptual errors in text-editor usage, in J. C. Chew and J. Whiteside (Eds.), *Proceedings of the CHI 90 Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, 1990.

Space Habitability



A three-dimensional interactive computer graphics package called PLAID is used to address human factors issues in spacecraft design and mission planning. Premission studies produced this PLAID rendition to show where an EVA astronaut would stand while restraining a satellite manually and what the IVA crewmember would be able to see from the window.

(See cover for the actual photo taken during mission from aft crew station.)